

# 数字人文“二手证据”循证范式研究： 以敦煌汉简为例

文玉锋<sup>1</sup>，马倩妮<sup>1</sup>，杨克虎<sup>2,3</sup>，周文杰<sup>1,2\*</sup>

(1. 西北师范大学 商学院，兰州 730070；2. 兰州大学 循证社会科学交叉创新实验室，兰州 730030；

3. 兰州大学 基础医学院循证医学中心，兰州 730030)

**摘要：**[目的 / 意义]随着数字人文研究范式的兴起，为研究者基于二手资源开展循证研究提供了契机。[方法 / 过程]本文首先梳理了数字人文“二手证据”循证范式相关的理论基础，并以敦煌汉简为例对数字人文“二手证据”循证范式进行了实证分析。通过来自 CNKI 的 287 篇敦煌汉简相关研究文献的编码分析，检验了信度和效度，并对类目体系展开了综合分析。[结果 / 结论]本文提出了循证数字人文基于“二手证据”展开循证研究的包括文献检索、类目编制、“二手证据”要素的综合分析 3 个阶段，以及证据检索、信度检验环节、效度检验环节和类目构建、证据要素综合 5 个环节构成的“三阶段五环节”范式。本研究所构建的范式，对循证数字人文学科、学术和话语体系建设具有一定的理论参照和实践指引价值。

**关键词：**循证数字人文；二手证据；“三阶段五环节”范式；敦煌汉简

**中图分类号：**G250.7

**文献标识码：**A

**文章编号：**1002-1248 (2022) 11-0026-12

**引用本文：**文玉锋，马倩妮，杨克虎，等. 数字人文“二手证据”循证范式研究：以敦煌汉简为例[J]. 农业图书情报学报，2022, 34(11): 26-37.

## 1 引言

循证数字人文的主要目标，是对蕴藏于数字化人文资源载体中丰富的证据要素加以挖掘，以备对这些隐性的证据要素加以高效率整理、高质量序化和高水

平复现。循证数字人文研究中的“二手证据”主要指研究者针对原始的记录资源，已展开相应的解析，从而获得的“次生证据”（例如，已发表的学术论文或其他研究成果）。在循证数字人文研究中，二手证据与一手证据的区别在于，二手证据并非天然地呈现在数字人文的载体（如文献资源）之中，而是经由研究者

收稿日期：2022-10-30

基金项目：国家社会科学基金重大项目“循证社会科学的理论体系、国际经验与中国路径研究”（19ZDA142）

作者简介：文玉锋，副教授，西北师范大学商学院。马倩妮，硕士研究生，西北师范大学商学院。杨克虎，教授，兰州大学循证社会科学交叉创新实验室、兰州大学基础医学院循证医学中心

\*通信作者：周文杰，教授，西北师范大学商学院，兰州大学循证社会科学交叉创新实验室，兼职教授，研究方向为信息社会问题。Email: wj\_lp@sina.com

对一手资料加以消化后获得的新的证据信息。

在中华民族五千年来源远流长的文化发展历程中, 卷帙浩繁的古籍文献承载着厚重的文化记忆与生生不息的民族精神, 是循证数字人文研究的重要素材库。过去一个多世纪的时间里, 在甘肃河西等地区发掘了大量以汉简为代表的简牍文献。据统计, 河西简牍占甘肃境内出土简牍的 90%, 占全国出土简牍的 1/4, 占全国汉简出土量的 80%以上<sup>[1]</sup>。这些宝贵的文化资源为循证数字人文研究者回溯两千多年前汉代文化气息, 复现汉朝社会风貌提供了关键素材。

迄今为止, 学者针对河西简牍展开了多维度的研究。在军事防御体系方面, 学者们基于河西汉简, 对汉代的多层次国防体系进行了研究<sup>[2]</sup>。也有学者以机构、制度、器物 3 个方面为切入点, 详细论述汉朝在河西的军事防御体系<sup>[3]</sup>。农业民生方面, 河西出土的汉简中有大量有关居民饮食生活的记载, 真实地反映了汉代河西居民的饮食文化<sup>[4]</sup>。在艺术方面, 研究者普遍认为, 河西简牍是汉简艺术宝库中的重要组成部分<sup>[5]</sup>。

敦煌汉简作为河西简牍中出土数量极多的简牍, 其研究成果也非常丰富。针对敦煌汉简研究的整理工作, 学者们从不同的角度进行了分类总结。有学者从述论考释类、语言文字类、人文地理类、书法艺术类、历谱历日类、文学类和医学类 7 部分入手, 对敦煌汉简的研究成果进行了整理<sup>[6]</sup>。也有学者对敦煌汉简的相关专著、语言文字文学、书法、年代考释等研究进行了整理论述<sup>[7]</sup>。针对敦煌汉简各个领域的研究成果也非常丰富。例如, 在语言文字方面, 不断有学者对敦煌汉简的简文进行释读、修改与补充<sup>[8,9]</sup>。在书法艺术方面, 有学者对敦煌汉简中的笔形、笔法进行研究, 探究了敦煌汉简的笔画变异路径<sup>[10]</sup>。在社会生活方面, 通过研究敦煌汉简中的私人书信, 研究者揭示了书信中反映出的时代背景<sup>[11]</sup>。此外, 也有学者提出敦煌汉简直接促成了近代简牍学的产生<sup>[12]</sup>。总的来看, 敦煌汉简的研究成果涉及多个方面, 对多个领域的研究都有着极高的史料价值。

本研究以河西简牍的典型代表——敦煌汉简相关研究成果为分析对象, 借助于质性分析的软件工具

Nvivo, 通过系统性编码, 对最近几十年来研究者针对敦煌汉简展开研究所获得的二手证据加以深度内容分析, 提取大量二手证据要素, 从而对数字人文研究中基于“二手证据”而展开循证研究的基本范式加以展示, 以便为循证数字人文学科、学术和话语体系建设提供理论参照和实践指引。

## 2 数字人文“二手证据”循证范式的理论基础

### 2.1 知识组织与数字人文的基础设施建设

数字人文横跨人文社科、计算机学科及图书馆职业等学科和职业领域, 其理论体系尚未发展成熟, 但从其演化路径观察, 整个数字人文的研究脉络正在不断地深化<sup>[13]</sup>。应用先进的信息技术对数字化的人文资源加以挖掘与复现, 堪称数字人文研究的基本特征。由于图书馆是文献资源的主要管理者, 因此, 刘炜等认为, 图书馆是数字人文不可或缺的基础设施<sup>[14]</sup>。刘芳等提出数字人文研究联盟多以高校和图书馆为依托, 涉及人文社会科学、数字艺术和其他相关领域, 为各类数字人文研究项目提供丰富的数据支撑、持续的资金支持、机构间的协同管理等<sup>[15]</sup>。

知识组织是大规模数字人文研究得以展开的前提。周文杰认为, 图书馆职业对于数字人文的参与, 主要体现在知识资源的整序和提供利用方面<sup>[16]</sup>。夏翠娟等则认为, 在参与数字人文研究方面, 图书馆职业的主要任务是利用新的技术手段重组资源, 为人文研究者提供更好的服务, 辅助人文研究, 而不是要深入某一个具体的领域, 代替人文研究者从事具体的研究工作<sup>[17,18]</sup>。王丽丽等<sup>[19]</sup>认为, 古籍知识关联是数字人文研究的关键视角。

图书馆职业如何保障数字人文呢? 首先, 图书馆学与数字人文所涉及的人文学科都共同关注基于序化的记录资源并以此来展开研究<sup>[20]</sup>。其次, 从知识资源的组织与使用的角度看, 数字人文不仅需要传统的检索与分析手段做出改变, 还需要通过关联分析、预测

分析以及非线性分析、量化比较、历史地图等方式，为人文研究带来更多方法的创新<sup>[21]</sup>。另外，图书馆职业关于知识资源的整序与数字人文之间的联系具体体现在：首先，经过图书馆职业整序的知识资源构成了数字人文研究的基础；其次，经过图书馆职业序化整理的知识资源不仅为开展数字人文提供了平台，也在很大程度上扩大了人文学者开展研究的样本范围<sup>[22]</sup>。刘炜认为，数字人文作为一个整体的跨学科研究领域，从传统的文献考据和现代的文献计量学中汲取了有益的方法论借鉴。而且，图书馆职业长期致力于发展的规范控制（特别是书目控制），借助于语义技术，可以为知识的形式化组织（采用本体技术）和知识服务提供了可信的编码基础，从而也为机器学习和人工智能的发展提供了宝贵的标注语料库。

## 2.2 循证效度协同

所谓循证效度协同，是指研究者借助循证的理念、工具与方法，通过对零散的定量或定性原始证据加以整合，从而有效提升其内外部效度，并以系统评价（Systematic Review）的形式对整合后的高效度证据加以报告的过程<sup>[23,24]</sup>。

在数字人文研究领域，研究者之所以能够基于循证方法而实现证据效度的协同，首先在于元分析等循证研究的工具使研究者可以通过提取具有高内部效度的研究证据，并将其加以综合（Research Synthesis），从而将多元化的数字人文研究情境也纳入了研究之中。具体而言，在数字人文研究中，针对定量成份比较高的研究发现，因为其内部效度较高而外部效度可能有所欠缺，因此，循证研究者通过将丰富的外部情境因素纳入研究，从而使数字人文研究成果实现内外部效度的协同。相反，针对质性成份比较高的数字人文研究成果，因其研究发现深刻植根于社会情境，因此具有较高的外部效度，但证据的内部效度则相对不足。为此，循证数字人文研究者基于更多个体证据的整合，实现内部效度的提升。总之，数字人文的发展历程表明，循证效度协同在数字人文研究领域不仅是可行性的，而且也是必要的。

## 3 研究设计

### 3.1 研究素材

数字人文研究展开的前提，是研究者广泛占有文献等记录资源为载体。本研究中，为充分展示数字人文研究者针对“二手证据”展开循证的范式，选定以敦煌汉简为主要分析对象。敦煌作为古代丝绸之路上的重镇，不仅拥有世界文化资源的宝库——莫高窟，而且也出土了大量汉简，是河西简牍中最重要的组成部分。1907年，英国人斯坦因于敦煌西北汉长城发掘出土705枚。1913—1915年斯坦因在敦煌、酒泉西北汉长城遗址发掘168枚。1920年，周炳南在敦煌小方盘城外采集17枚。1944年，夏鼐在敦煌小方盘城汉长城遗址发掘48枚。1977年，嘉峪关文管所在玉门花海汉长城遗址掘得91枚。1979年，甘肃省文物工作队在敦煌马圈湾发掘出土1217枚。1981年，敦煌市博物馆在敦煌酥油滩掘得76枚。1987—1988年，敦煌市博物馆在汉代烽燧遗址时采集137枚。1990—1992年，甘肃省文物考古研究所在敦煌悬泉置遗址发掘出土汉代简牍20000余枚。到目前为止，敦煌及周边区域的汉简发掘工作仍在继续。

敦煌汉简大部分是原始的档案文书和实录，对于研究古代政治、经济、军事、法律、文化和科技堪称现存史料中最详尽、最真实可靠的第一手资料，具有重要的史料价值和学术价值。20世纪以来，敦煌汉简研究广为学界所重视，围绕包括敦煌汉简在内的河西简牍展开的学术研究也已产出了大量成果。本研究旨在以20世纪以来学术界围绕敦煌汉简展开的研究所取得的学术成果为研究对象，应用内容分析的方法，对敦煌汉简研究的历史脉络做出解析，并在揭示现有研究特征的基础上，对敦煌汉简研究的未来趋向做出预测。

### 3.2 数据来源

2022年11月23日，在中国知网中，对“篇关摘”（篇名、关键词、摘要）中含有“敦煌汉简”的文献进行检索，共得到643条检索记录。将这些检索记录导



人文献管理软件 Endnote 后, 先进行去重处理, 剔除重复的检索结果 4 条。进而, 去除了发表于报纸的新闻报道 6 篇, 并删除元数据存在期刊、作者等重要缺项的数据 130 条。经过上述步骤的数据清洗, 本研究最终获得有效的分析数据 503 条。为确保数据质量, 对已清洗完成的数据, 由两名研究者分别逐篇阅读文章摘要, 对文章主题与“敦煌汉简”的关联程度做出判断。最终, 剔除关联度不高的文献 216 篇。经过上述步骤, 本研究最终获取用于展开研究的学术论文 287 篇。

### 3.3 分析方法

本研究主要采用内容分析法。内容分析法是一种针对文本内容进行深度解析的研究方法。内容分析法 (Content Analysis) 是一种以文献的内容特征为统计单元, 对文本中的主题内容进行客观、系统、定量分析, 以揭示文本中所蕴含主题的数量关系和内在规律的竞争情报分析方法。内容分析法起源于 20 世纪初期的传播学领域, 被认为是对大众传媒进行研究的有效手段之一。早在第二次世界大战期间, 内容分析法就在情报战中的军事情报研究中得到了广泛应用。二战以后, 新闻传播学、政治学、图书馆学、社会学等领域的专家学者与军事情报机构一起, 对内容分析方法进行了多学科研究, 使其应用范围大为拓展。

美国学者 B. Bereson 出版于 1952 年的著作《内容分析: 传播研究的一种工具》标志着内容分析作为一种文本材料分析方法正式得到了确认。简单而言, 内容分析法是一种对具有明确特征的传播内容进行的客观、系统和定量的描述的研究技术, 该方法具有客观性、系统性、非接触性、相对主观性等诸多特征和优势。

### 3.4 分析工具

本研究采用质性分析软件工具 Nvivo, 以提高内容分析的效率和科学性。本研究应用 Nvivo 开展敦煌汉简相关文献内容分析的具体操作过程是: 首先, 两名研究者“背靠背”逐篇阅读所检索到的文献全文, 分别提取类目, 并将其录入 Nvivo; 其次, 对两名研究者

提取的类目进行了一致性分析, 计算了信度系数 (信度分析的具体结果请见本文第三部分); 第三, 在 Nvivo 中, 对各类目之间的关联关系进行了建模, 以备后续分析。

### 3.5 研究过程

如前文所述, 本研究首先在 CNKI 中针对敦煌汉简研究相关文献进行了系统性检索, 并应用 Nvivo 软件进行了类目构建与类目间关系的分析。本研究的具体步骤是: ①两位编码员分别进行了类目提取; ②基于前 20 篇文献进行了编码一致性分析; ③继续拓展一致性分析范围, 逐步涵盖全部文献; ④对所构建的类目体系进行了信度和效度检验; ⑤在构建了类目体系的基础上, 就类目间关系进行了综合分析。

## 4 研究结果

### 4.1 数字人文“二手证据”循证范式的信度基础

定性研究目前主要采用两种编码方式: 一种是完全开放式的编码, 另外一种严格按照已有理论设定编码变量。本研究是在熟读材料形成初步理论框架的基础上展开编码的, 在编码过程中若发现新编码的变量就将其添加到框架中。具体而言: 泛读 287 篇有关敦煌汉简的文章并大致确定一级维度, 也就是树状节点, 进而形成研究框架, 再由两名研究者对文献信息进行独立编码, 形成若干子节点, 最后通过合并同类项等, 将子节点放至相应的树状节点下。

为保证编码的一致性, 提高质性研究的信度, 通常要从样本中随机抽取 10% 至 15% 的数据展开分析。本研究从 287 篇文章中随机抽取了 40 篇文章分两轮进行信度检验, 具体过程如下。

第一轮: 从随机抽取的 40 篇文献中任意选取 20 篇文章让两位研究者进行独立编码, 并计算信度。背靠背编码完成后, 直接进行编码比较, 此时的信度系数为 0.66, 随后, 两位研究者仔细调查存在差异的编码, 结果发现军事战争和地理考释这两部分的编码差

异最大，经过斟酌研讨后达成一致意见并再次对原来的 20 篇文章进行编码，此时信度系数达到 0.853，具体结果如表 1 所示。

$$\text{信度} = \frac{2M}{N_1 + N_2}$$

上式中， $M$  是编码决策过程中，两个编码员意见一致的数量。 $N_1$  和  $N_2$  是两个编码员各自进行的编码决策的总量。通过上述计算公式，我们可以得到第一轮信度系数为  $=2*272/(316+322)=0.853$ 。由于两位研究者之间的信度高于 0.8，说明两位研究者编码基本一致，研究是可靠的。

第二轮：对余下的 20 篇文献，继续由两位研究者进行第二轮独立编码，并计算信度，结果如表 2 所示。

由第二轮编码的一致性分析结果可见，第二轮类目提取的信度为 0.862。总之，两轮类目提取结果均表明，本研究关于“二手证据”类目分析的结果是可靠的。

4.2 类目的效度

本研究对于类目效度的分析主要采用专家判断法，具体过程是：首先，类目的选择经过两名研究者反复的研讨，对于不能达成一致意见的类目，咨询在该领域进行深度研究的权威专家意见；其次，两名研究者

表 1 第一轮编码的一致性

Table 1 Consistency of the first round of coding

序号	编码员一	编码员二	总计一致参考点	序号	编码员一	编码员二	总计一致参考点
1	11	8	8	17	10	12	10
4	12	11	11	18	11	10	10
5	5	7	4	23	6	5	5
6	15	15	12	43	2	6	1
7	7	8	6	46	40	37	37
9	18	20	14	50	16	15	13
10	13	13	10	70	32	42	31
11	28	30	28	82	38	30	25
12	23	24	23	190	3	4	2
13	26	25	22	总计	316	322	272

表 2 第二轮编码的一致性

Table 2 Consistency of the second round of coding

序号	编码员一	编码员二	总计一致参考点	序号	编码员一	编码员二	总计一致参考点
117	13	10	9	195	25	24	24
122	2	1	1	196	19	18	18
127	3	3	3	210	3	2	2
131	14	14	14	212	7	5	4
147	19	18	16	213	21	20	17
151	12	9	9	214	16	14	11
171	6	4	4	247	29	30	29
172	18	18	17	307	14	13	11
179	2	5	2	344	18	17	10
47	6	6	6	372	7	9	6
合计	254	240	213				

事先进行了相关培训，在整个编码过程中严格遵守相关编码规则与程序。

4.3 类目体系与信息提取的单元

分析单元是内容分析法中最小的元素，它可以是段、句子，也可以是字、词等。本研究是将敦煌汉简相关文献的关键语句作为分析单元，根据已经构建的七维分析框架，将分析类目确定为如表 3 所示的类目体系。

如表 3 所示，本研究共设置了七大主节点，并设置了与此相对应的 30 个子节点，将敦煌汉简相关的 287 篇文献编码到 Nvivo 中，共得到了 5 181 个参考点。如图 1 所示，从宏观角度来讲，语言文字文学研究这个维度参考点占比最高，达到了 42%，人文地理维度参考点占比也较高，达到了 34%，中西交流这个维度参考点为第三高，达到了 9%，相较而言，敦煌汉简所涉及的法律维度的内容较少，参考点占比仅为 4%，不难看出，敦煌汉简所记录的内容涉及方方面面，人们也越来越关注到敦煌汉简所蕴含的史学价值，不断解读敦煌汉简以期挖掘到两汉时期的风情地貌等

优质信息。

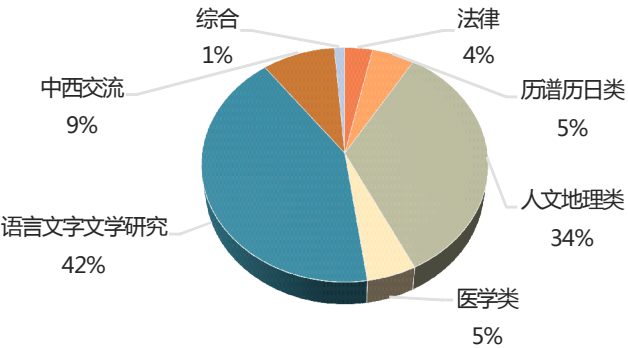


图 1 7 个类目证据节点占比

Fig.1 Proportion of nodes in seven evidence categories

4.4 循证数字人文研究中“二手证据”要素的提取

如上文所述，通过系统性的循证方法，在确保证据获取信度和效度的基础上，本研究初步构建了敦煌汉简研究的二手证据类目体系。进而，按照循证数字人文研究的规范流程，本团队展开了证据要素的提取工作。经过两位编码员的共同努力，最终在 287 篇敦煌研究学术文献中提取了 5 181 个全证据参考点。具体如表 4 所示。

表 3 类目体系表

Table 3 Categorical system

一级节点	二级节点	一级节点	二级节点
医学类	药、药函及制度	法律	刑罚
	医案与方剂		律、令、品及司法文书
	疾病类型	中西交流	丝绸之路
	法医验视		西域交往
	兽医资料		其他研究
语言文字文学研究	卜书	人文地理类	政治
	书法		军事
	释读及校订		经济
	文化典籍		农业民生
	书信		地理考释
	姓氏人名研究	历谱历日类	时间考释
	传说故事		术数研究
	字形及语法研究	综合类	相马术
	简牍学		敦煌郡
	研究论著目录		著作之评价

chinaXiv:202303.10378v1

chinaXiv:202303.10378v1

表 4 敦煌汉简研究中的人文地理类“二手证据”参考点 单位：个

Table 4 Reference points of Second-hand Evidence in human geography in the study of Dunhuang bamboo slips in Han Dynasty								
名称	文件	参考点	名称	文件	参考点	名称	文件	参考点
政治	52	803	户籍民	1	17	诣官符	1	3
邮驿制度	29	301	政治文书	15	112	经济	16	189
邮驿传置	25	218	诏书	8	60	物价	5	77
车马	9	63	“传信”	2	13	会计记账	6	49
基层组织、里、亭、乡	8	166	“刺”	4	10	西域贸易	2	25
“亭”的研究	1	42	“移”	1	10	契约	2	24
治安之亭	1	29	失亡传信册	1	9	雇佣券	1	14
侯望之亭	1	13	出入关致	1	3	农业民生	22	184
悬泉置	4	37	简牍档案	1	3	农田水利	8	58
邑名	1	34	举书	1	3	牲畜研究	6	50
基层组织“乡”	2	21	惊备檄	1	1	马食	4	41
效谷县	1	21	军事	30	407	牛和牛车	2	9
基层组织“里”	1	11	军队战争	7	192	人的传食	2	32
官阶及相关	10	151	王莽西域战争	5	148	农业生产	2	17
官吏	7	110	汉匈车师之争	2	19	度量单位	2	14
敦煌太守	2	56	汉匈乌孙之争	1	17	“菴”	1	9
“中舍”考	1	13	戍卒赴边	7	66	大小斛石	1	5
敦煌郡诸侯官	1	12	军事制度	8	48	动物皮革	1	7
护羌使者	1	9	戊己校尉	4	41	社与社祭	1	3
侯官	1	5	中垒校尉	2	5	体育	1	3
译官	1	4	边塞挈令	1	4	地理考释	22	175
官守候	1	2	都尉	2	3	玉门关关址	15	121
护羌校尉	1	1	军事机构	1	2	悬泉置地理	3	15
职官体系	3	29	武器装备	4	36	三节、泉都	1	13
秩级“真二千石”	1	12	烽火品约中的蓬与其他	1	23	文德	2	12
政治事件	2	32	防御体系	2	21	大煎都管辖	1	9
逐捕辛興	1	21	信符	1	21	池头县	1	1
王莽改制	1	11	巡迹符	1	11	效谷县	1	1
“適”令	1	23	出入关符	1	3	骊靳县	1	1

如表 4 所示，在 7 个维度上都分布了大量的证据信息点。在循证数字人文研究中，这些证据参考点为研究者开展深入的数字人文研究提供了广阔的空间。

4.5 循证数字人文研究中“二手证据”要素的综合分析

本研究通过系统性编码，将敦煌汉简研究中所获取的二手证据编码为 7 个维度。这 7 个维度在总体二

手证据体系中的分布具有疏密有致的特征。图 2 展示了在 7 个维度各子主题上，敦煌研究相关二手证据的分布状况。图 2 中，面积越大，表示该类二手证据占比越高；内部结构分布划分的小方块越多，则表示该维度下的二手证据越多元。

以 7 个维度中的“人文地理维度”为例，该维度下的二手证据又可具体分为地理考释、经济、军事、农业民生、政治等主题。这些具体在此维度下的分布

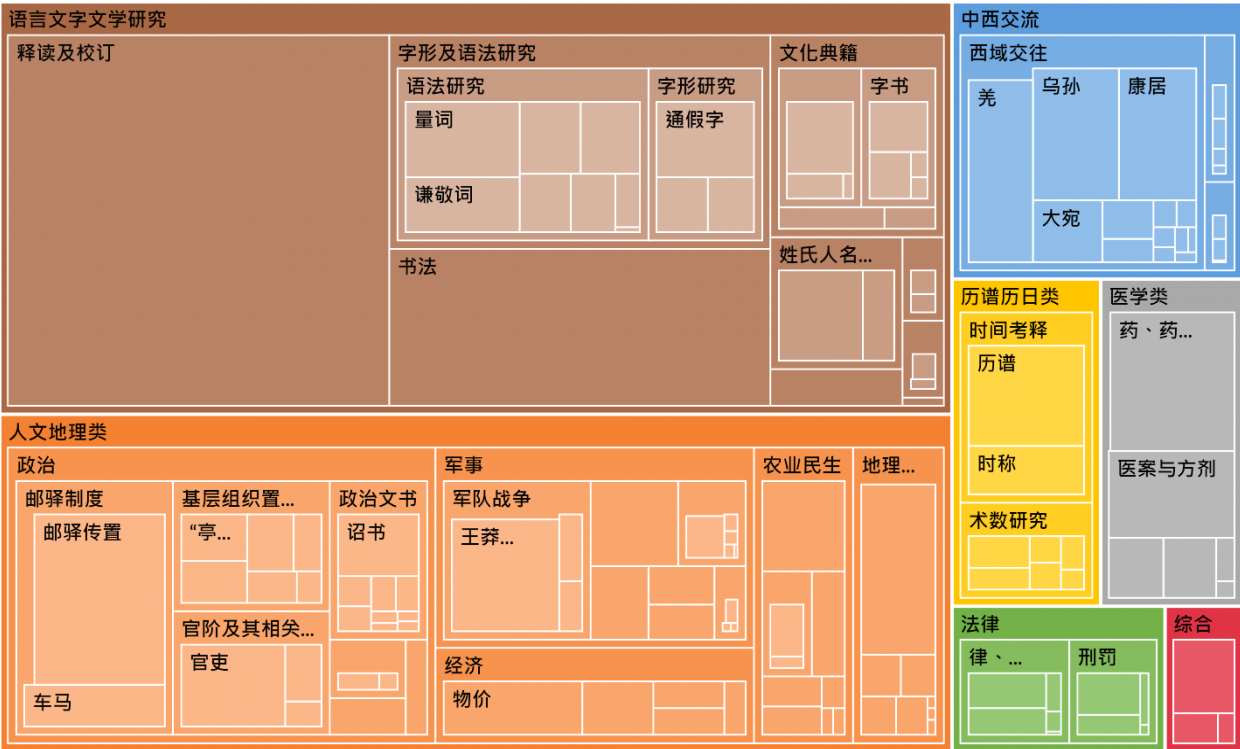


图 2 7个维度下二手证据的主题分布

Fig.2 Subject distribution of second-hand evidence in seven dimensions

如图 3 所示。

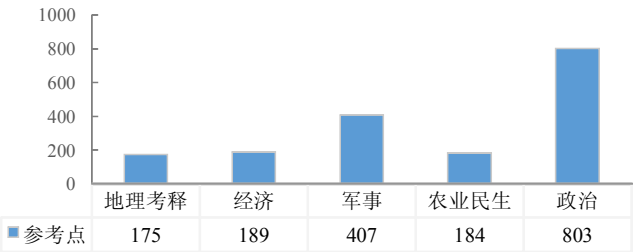


图 3 人文地理维度涉及的主要二手证据类型

Fig.3 Main second-hand evidence types involved in human geography dimension

表 5 进一步展开了人文地理维度下具体类目的参考点。从微观角度来说，人文地理维度囊括政治、经济、军事、农业民生和地理考释，内涵丰富。

表 5 中，玉门（玉门关）出现频次高达 190 次，阳关出现 28 次，可见随着汉简的不断出土，勘探技术的日益精湛使得玉门关关址、阳关关址的考察成为了近些年的研究热点。

苜蓿出现 82 次，粮食出现 26 次，可见中国以农

表 5 人文地理维度下的具体二手证据参考点

Table 5 Specific second-hand evidence reference points under the human geography dimension

“人文地理”相关关键词和词频/次		
敦煌（336）	汉简（213）	玉门（玉门关）（190）
西域（114）	苜蓿（82）	文书（67）
使者（41）	机构（37）	匈奴（36）
接待（33）	战争（31）	军事（28）
阳关（28）	粮食（26）	官吏（25）

业经济为主的特征，除了对人食的研究，还聚焦于马食的研究，可见马在汉代的重要地位，不仅仅是牲畜更是关乎战斗硬实力的驱动因素之一。

西域词频为 114 次，使者词频为 41 次，机构出现 37 次，接待词频为 33 次，其中机构多指代的是“置”，文献主要论述了悬泉置的功能和级别，让我们了解了汉代邮驿系统一级机构的功能和作用；西域词频之高显示了汉王朝对西域关系的重视，西域各国也为维系与汉王朝的友好关系，频繁派遣各国使者前往汉王朝，



其中涉及西域诸国，如车师、大宛、小宛、于阗、康居、乌孙等；官吏出现 25 次，这些文献主要论述了两汉时期的职官制度，例如探讨了敦煌郡属县长官“令”“长”的变化以及郡县吏员的守官制度，也让我们对“真二千石”等秩级有了明晰的认识，同时探讨了属官问题，例如对戊己校尉的设置年代、隶属关系等做了剖析；文书出现频次也很高，主要探讨的是各种政治文书，展现了当时官文书的用语制度以及千姿百态的政治文书的类型如：传信、刺、出入关致、惊备缴等。

匈奴出现 36 次，军事 28 次，战争 31 次，详细记录了汉朝与匈奴关系的变化，由最开始的和亲政策到后续战争爆发，足见汉朝威慑力的降低，让我们对王莽西域战争具体战况、失利原因更加明白。

数字人文研究者也可以对二手证据展开各种类型的交叉分析。本研究中，我们首先将检索到的全部文章按照研究的汉简类型进行分类，从研究数量来看，大部分对敦煌汉简的研究并不是针对某一种汉简进行的分析（图 4）。其中，对于整体敦煌汉简的研究有 163 篇，占比最高。除此之外，对于悬泉汉简的研究也比较多，有 81 篇文章针对敦煌汉简中在悬泉置出土的悬泉汉简进行了研究。最后，有 34 篇文章针对出土于马圈湾的敦煌汉简进行了研究。

如表 6 所示，在关于敦煌汉简研究的 163 篇文章中，对于语言文字文学的分析内容是最多的；在关于悬泉汉简的 81 篇研究中，对于人文地理的研究是最多的；而在关于马圈湾汉简的 34 篇研究中，则对于语言

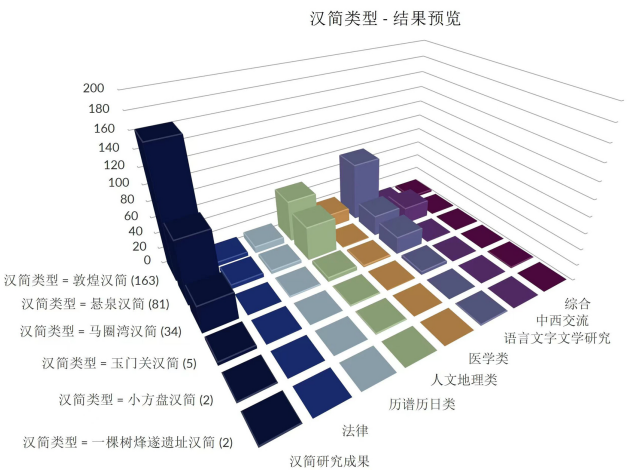


图 4 不同出土地敦煌汉简研究主题的交叉图

Fig.4 Cross themes analysis of Dunhuang bamboo slips in Han Dynasty unearthed in different places

文字文学的研究最多。

5 小 结

5.1 数字人文“二手证据”循证的“三阶段五环节”范式

本研究以敦煌汉简为例，通过系统性编码，全面展示了循证数据人文研究中，如何针对二手证据展开研究的过程。概括而言，循证数字人文基于“二手证据”的循证范式形成“三阶段五环节”的研究范式。

数字人文研究者基于“二手证据”展开循证研究的第一个阶段，是展开广泛文献检索，获取尽可能全

表 6 不同出土地汉简研究主题的分布 单位：篇

Table 6 Distribution of research topics of Dunhuang bamboo slips in Han Dynasty in different excavations							
汉简类型	敦煌汉简	悬泉汉简	马圈湾汉简	玉门关汉简	小方盘汉简	一棵树烽隧遗址汉简	合计
法律	6	6	0	1	0	0	13
历谱历日类	9	5	2	0	0	0	16
人文地理类	54	44	6	2	2	0	108
医学类	18	1	3	1	0	0	23
语言文字文学	76	27	23	5	0	2	133
中西交流	5	23	1	0	0	0	29
综合	5	2	1	0	2	0	8
合计	163	81	34	5	2	2	285

面的研究文献, 以便为展开“二手证据”的分类与提取奠定基础。

基于信度、效度等科学性评价指标, 对上一阶段所获得的“二手证据”载体展开类目编制, 并提取证据要素, 是数字人文研究者展开“二手证据”循证研究的第二个阶段。这一阶段具体可分解为3个环节, 即信度检验环节、效度检验环节和类目构建环节。

循证数字人文研究者基于“二手证据”展开研究的最后一个阶段, 是基于“二手证据”要素的综合分析。如上文所述, 研究者既可针对证据要素展开单维度的分析, 又可以综合多种证据要素, 展开系统性的综合解析。

## 5.2 后续研究

百年以来, 河西简牍已成为简牍学研究的重要组成部分, 而敦煌汉简是其中不可或缺的重要组成部分。本研究以敦煌汉简这一宝贵的文化遗产为对象, 就循证数字人文研究中基于“二手证据”展开研究的范式进行了深入讨论。本研究所确立的“三阶段五环节”的研究范式, 为研究者充分利用学术文献等载体中所蕴含的丰富“二手证据”元素, 有效展开循证数字人文提供了启示。然而, 循证数字人文研究是一个新兴的研究领域。在后续的研究中, 还需要根据循证社会科学的一般规程, 从更多维度、基于更先进的技术方法展开深入研究。例如, 在本研究所涉及的敦煌汉简的相关研究梳理中可以看出, 数字人文研究的基础设施尚需要大力完善。具体而言, 基于敦煌汉简而展开的数字人文研究亟待采取新科技手段, 适应数智化时代的需求, 争取为学界提供更便捷的获取方式、更全面的文献著录信息和更高质量的内容, 从而从多重证据挖掘的角度提供智慧化服务, 使宝贵的敦煌汉简资源发挥出更大的学术价值, 以促进学术研究的繁荣。

### 参考文献:

- [1] 孙占鳌. 河西简牍学百年发展述论[J]. 丝绸之路, 2015(4): 15-20.  
SUN Z A. On the development of Hexi Jiandu in the past century[J]. The silk road, 2015(4): 15-20.

- [2] 张瑛. 从出土简牍看汉王朝的河西军事防御[J]. 甘肃社会科学, 2016(6): 94-99.  
ZHANG YING. Military defense of Hexi in Han dynasty from unearthed Jiandu[J]. Gansu social sciences, 2016(6): 94-99.
- [3] 尹亮. 从河西简牍看汉代对匈奴的军事防御体系[D]. 兰州: 兰州大学, 2016.  
YIN L. The research on the Han dynasty military defense system to the Huns through Hexi Jiandu[D]. Lanzhou: Lanzhou university, 2016.
- [4] 孙占鳌, 刘生平. 从出土简牍看汉代河西饮食[J]. 甘肃社会科学, 2014(6): 90-94.  
SUN Z A, LIU S P. Hexi diet in Han dynasty viewed from unearthed Jiandu[J]. Gansu social sciences, 2014(6): 90-94.
- [5] 俞治. 浅谈河西简牍对当代书法创作的借鉴意义[J]. 中国民族博览, 2019(3): 110-111.  
YU Z. On the significance of Hexi Jiandu for contemporary calligraphy creation[J]. Panorama of Chinese nationalities, 2019(3): 110-111.
- [6] 潘芳. 近20年来敦煌汉简研究综述[J]. 西安文理学院学报(社会科学版), 2015, 18(5): 35-40.  
PAN F. Review of researches on Dunhuang bamboo slips in the past 20 years[J]. Journal of Xi'an university (social sciences edition), 2015, 18(5): 35-40.
- [7] 裴永亮. 近二十年来敦煌汉简研究综述[J]. 管子学刊, 2015(2): 121-126.  
PEI Y L. A summary of the studies on Dunhuang Hanjian in the past 20 years[J]. Guanzi journal, 2015(2): 121-126.
- [8] 张俊民. 斯坦因所获敦煌汉简释读札记[J]. 秦汉研究, 2022(1): 232-244.  
ZHANG J M. Reading notes on Dunhuang Hanjian obtained by stein[J]. Qin and Han researches, 2022(1): 232-244.
- [9] 郑雯, 刘向培. “率”在敦煌汉简中的分指代词用法新探[J]. 现代语文, 2022(8): 17-21.  
ZHENG W, LIU X P. A new probe into the use of discriminative pronouns of "lv(率)" in Han bamboo slips of Dunhuang[J]. Modern Chinese, 2022(8): 17-21.
- [10] 雷黎明, 王欢. 汉隶到章草笔画变异现象探略——以敦煌汉简为例[J]. 档案, 2021(11): 24-29.  
LEI L M, WANG H. A brief study on the variation of the strokes from

Hanli to Zhangcao: Taking Dunhuang Hanjian as an example[J]. Archives, 2021(11): 24-29.

[11] 郑芙玉. 从《敦煌汉简》看私人书信[J]. 大众文艺, 2021(1): 11-12.

ZHENG F Y. Private letters from Dunhuang Hanjian[J]. Popular literature and art, 2021(1): 11-12.

[12] 常燕娜. 敦煌汉简与简牍学的产生及新发展[J]. 丝绸之路, 2022(2): 28-31.

CHANG Y N. The birth and new development of Dunhuang Hanjian and Jianduxue[J]. The silk road, 2022(2): 28-31.

[13] 许鑫, 陈路遥, 杨佳颖. 数字人文研究领域的知识网络演化——基于题录信息和引文上下文的关键词共词分析[J]. 情报学报, 2019, 38(3): 322-334.

XU X, CHEN L Y, YANG J Y. Knowledge network evolvement in digital humanities domain: Based on keywords co-word analysis of bibliography and citation context[J]. Journal of the China society for scientific and technical information, 2019, 38(3): 322-334.

[14] 刘炜. 作为数字人文基础设施的图书馆: 从不可或缺到无可替代[J]. 图书馆论坛, 2020, 40(5): 1-2

LIU W. Library as digital human infrastructure: From indispensable to irreplaceable[J]. Library tribune, 2020, 40(5): 1-2.

[15] 刘芳, 余望枝. 数字人文环境下图书馆资源建设与服务模式研究[J]. 图书馆, 2022(4): 42-49.

LIU F, YU W Z. Research on library resource construction and service mode under digital humanistic environment[J]. Library, 2022(4): 42-49.

[16] 周文杰. 图档档学科发展的数据化趋向解析——基于 2019 年度学术热点的系统性文献调查[J]. 情报资料工作, 2020, 41(2): 20-30.

ZHOU W J. An analysis of the datafication trends of the discipline development of library, information and archives: Literature survey based on academic hotspots in 2019[J]. Information and documentation services, 2020, 41(2): 20-30.

[17] 夏翠娟. 面向人文研究的“数据基础设施”建设——试论图书馆学对数字人文的方法论贡献[J]. 中国图书馆学报, 2020, 46(3): 24-37.

XIA C J. The construction of "data infrastructure" for humanities research: the methodological contribution of library science to digital humanities[J]. Journal of library science in China, 2020, 46(3): 24-37.

[18] 夏翠娟, 贺晨芝, 刘倩倩, 等. 数字人文环境下历史文献资源共建共享模式新探[J]. 图书与情报, 2021(1): 53-61.

XIA C J, HE C Z, LIU Q Q, et al. New exploration on the co-construction and sharing mode of historical literature resources in the digital human environment[J]. Library & information, 2021(1): 53-61.

[19] 王丽丽, 张宁. 数字人文视角下的古籍知识关联探析[J]. 农业图书情报学报, 2022, 34(9): 51-59.

WANG L L, ZHANG N. Knowledge correlation of Chinese ancient books from the perspective of digital humanities[J]. Journal of library and information science in agriculture, 2022, 34(9): 51-59.

[20] KOLTAY T. Library and information science and the digital humanities: Perceived and real strengths and weaknesses[J]. Journal of documentation, 2016(4): 781-792.

[21] 韩豫哲. 国外数字人文项目特点分析及启示[J]. 图书馆, 2019(11): 18-23.

HAN Y Z. Analysis and enlightenment of the characteristics of foreign digital humanities projects[J]. Library, 2019(11): 18-23.

[22] 朱本军, 聂华. 跨界与融合: 全球视野下的数字人文——首届北京大学“数字人文论坛”会议综述[J]. 大学图书馆学报, 2016(5): 16-21.

ZHU B J, NIE H. Crossing boundaries and engaging communities: Digital humanities in a global perspective - Summary of the first Peking university "digital humanities forum" conference[J]. Journal of academic library, 2016(5): 16-21.

[23] 周文杰, 赵悦言, 魏志鹏, 等. 循证视角下文献证据检索的科学性评价: 缘起、指标与趋势[J]. 图书与情报, 2021(6): 31-36.

ZHOU W J, ZHAO Y Y, WEI Z P, et al. Scientific evaluation of literature evidence retrieval quality from the perspective of evidence-based research: Initiation, index and trend[J]. Library & information, 2021(6): 31-36.

## Exploring the Paradigm of Secondhand Evidence Based Digital Humanity Research: A Case Study of Dunhuang Wooden and Bamboo Slips of Han Dynasty

WEN Yufeng<sup>1</sup>, MA Qianni<sup>1</sup>, YANG Kehu<sup>2,3</sup>, ZHOU Wenjie<sup>1,2\*</sup>

(1. Business School of Norwest Normal University, Lanzhou 730070; 2. Cross-innovation Laboratory of Evidence-based Social Science of Lanzhou University, Lanzhou 730030; 3. Evidence-based Medical Center of School of Basic Medical Sciences of Lanzhou University, Lanzhou 730030)

**Abstract:** [Purpose/Significance] The research paradigm of digital humanities provides an opportunity for researchers to conduct evidence-based research based on second-hand resources. The second-hand evidence in the evidence-based digital humanities research mainly refers to the secondary evidence obtained by researchers through analysis of the original literature records. Academic papers are such an example of second-hand evidence and analysis of these papers enables us to be focused on the exploration, discussion and comprehension of the specific evidence. [Method/Process] This paper aims to shed light on the theoretical basis related to second-hand evidence based digital humanities research, and conduct an empirical analysis with Dunhuang wooden and bamboo slips of Han Dynasty. Through the coding analysis of 287 relevant research papers downloaded from the CNKI database, the reliability and validity of the research were tested, and a comprehensive analysis of the category system was conducted. On this basis, two independent coders systematically coded all academic papers by, and a coding system was formed with reliable reliability and validity. This coding system provided us a foundation to summarize the procedure of extracting the second-hand evidence from academic papers in the digital humanities research. [Results/Conclusions] This paper issued a three-stage and five-link procedure for second-hand evidence based digital humanity research. The "three stages" refer to the systematical literature retrieval, category formulation and comprehensive analysis of the second-hand evidence elements and the five links include evidence retrieval, reliability test, validity test, category construction and evidence element synthesis. The procedure developed by this research is expected to shed light on the knowledge organization of ancient literature works and lay a foundation for the digital humanities research based on the secondary resources like academic papers. Thus, based on the above procedure, the researchers in the digital humanities area are able to extract fruitful second-hand evidence from secondary resources (academic papers, for example) instead of acquiring initial evidence. This procedure may benefit the evidence-based digital humanity research because it is always a challenging task for individual researcher to access the initial research material like Dunhuang wooden and bamboo slips of Han Dynasty. In summary, the paradigm constructed in this study has certain theoretical reference and practical guidance value for the construction of evidence-based digital humanities subject, and academic and discourse system. Thus, it is expected to benefit the construction of independent knowledge system of evidence-based digital humanities with Chinese characteristics.

**Keywords:** evidence-based digital humanity; second hand evidence; three stages and five links paradigm; Dunhuang wooden and bamboo slips of Han Dynasty